ARI Research Note 97-28

Examining the Effects of Cognitive Consistency Between Training and Displays

Leonard Adelman
George Mason University

Matthew Christian
System Planning Corporation

James Gualtieri
Enzian Technology, Inc.

Karen Johnson
George Mason University

Research and Advanced Concepts Office
Michael Drillings, Chief

August 1997

United States Army
Research Institute for the Behavioral and Social Sciences

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency Under the Jurisdiction of the Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Director

Research accomplished under contract
for the Department of the Army

George Mason University

Technical review by

Michael Drillings

NOTICES

DISTRIBUTION: This report has been cleared for release to the Defense Technical Information Center (DTIC) to comply with regulatory requirements. It has been given no primary distribution other than to DTIC and will be available only through DTIC or the National Technical Information Service (NTIS).

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The views, opinions, and findings in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED (from. . . to)
1997, August	Final	June 1992-January 1996

4. TITLE AND SUBTITLE
Examining the Effects of Cognitive Consistency Between Training and Displays

5a. CONTRACT OR GRANT NUMBER
MDA903-92-K-0134

5b. PROGRAM ELEMENT NUMBER
0601102A

6. AUTHOR(S)

Leonard Adelman (George Mason University), Matthew Christian (System Planning Corp.), James Gualtieri (~~Exian~~ Technology, Inc.), and Karen Johnson (George Mason University) Enzian

5c. PROJECT NUMBER
B74F

5d. TASK NUMBER
2901

5e. WORK UNIT NUMBER
C01

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Dr. Leonard Adelman
c/o Department of Operations Research
George Mason University
Fairfax, VA 22030-4444

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Institute for the Behavioral and Social Sciences
ATTN: PERI-BR
5001 Eisenhower Avenue
Alexandria, VA 22333-5600

10. MONITOR ACRONYM
ARI

11. MONITOR REPORT NUMBER
Research Note 97-28

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
COR: Michael Drillings

14. ABSTRACT *(Maximum 200 words)*:

This paper describes the third and final experiment performed on Contract MDA903-92-K-0134. This experiment tested the "display cognitive consistency hypothesis" proposed in Adelman, Bresnick, Black, Marvin, and Sak (in press). This hypothesis states that the effectiveness of a display format for decision aiding systems, like Patriot, depends on the consistency between how the system displays its reasoning process and how the person is processing the information. Results of an experiment using a simulated Army air defense task and college students found support for the hypothesis, but only at a situation-specific, not global, level. Although unexpected, these results were consistent with other research performed on this contract, indicating the importance of situation-specific context for understanding judgment and decision processes in individual and group settings.

15. SUBJECT TERMS
Judgment heuristics; Cognitive biases; Cognitive engineering; Army air defense; Display design

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON (Name and Telephone Number)
16. REPORT	17. ABSTRACT	18. THIS PAGE			
Unclassified	Unclassified	Unclassified	Unlimited	67	

EXAMINING THE EFFECTS OF COGNITIVE CONSISTENCY BETWEEN TRAINING AND DISPLAYS

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Preceding Page Blank

Page

LIST OF FIGURES (Continued)

EXAMINING THE EFFECTS OF COGNITIVE CONSISTENCY BETWEEN TRAINING AND DISPLAYS

INTRODUCTION

Substantial research (e.g., see Heath, Tindale, Edwards, Posavac, Bryant, Henderson-King, Suarez-Balcazar, & Myers, 1993) has demonstrated that people use judgment heuristics to process information in many different domains, and that these heuristics can sometimes result in biased judments. Our concern has been the use of judgment heuristics in settings where Army personnel use computer systems to assist them in making judgments and decisions. In particular, we have performed a series of experiments with Army air defense personnel directed at understanding (1) what contextual features cause individual operators to use different heuristics; (2) whether teams also use these heuristics; and (3) whether training programs and computer displays could mitigate the use of heuristics, and their potentially negative effect under certain circumstances.

Specifically, research conducted by Adelman and Bresnick (1992), Adelman, Bresnick, and Tolcott (1993), and Adelman, Bresnick, Black, Marvin, and Sak (in press) on an earlier contract studied the effects of information order on the decisions of Army air defense personnel. These studies found that trained air defense personnel, when individually evaluating unknown aircraft tracks on the Patriot air defense simulator, made significantly different probability estimates (and identification and engagement decisions) when the same, conflicting information was presented in two different ordered sequences. This was referred to as an order effects bias.

The decisions were considered biased because (1) there was no empirical data indicating that different ordered sequences should result in different decisions, and (2) evidence indicating that it should not. Both training protocol and the recommendation algorithm in the Patriot system indicate that personnel should use an additive rule to process conflicting information about an aircraft track. An additive rule will lead to the same decision regardless of the order in which the same information is received by the operator. It will not result in order effects.

The goal of the current contract was to extend the research on cognitive heuristics and biases from individual to team decision making. Toward that end, we performed three experiments on the current contract. The first experiment studied whether the order effect bias also existed for Patriot teams. We found that it did, depending on how the aircraft looked on the Patriot display and the prior information about it. The second experiment studied whether training designed to increase communication among Patriot team members reduced the size of the order effect, thereby improved Patriot team performance. We found that it did not. The third experiment, which is described herein, studied whether features of the computer display could reduce order effects. That experiment, which was with college students, found that it could (or couldn't) depending on the situation-specific characteristics of the display, the track, and the operators' training. Before describing this experiment, we review the first two experiments performed on the contract.

In the first experiment, Adelman, Bresnick, Christian, Gualtieri, and Minionis (1995a) found that, on the average, the decisions of Patriot teams depended on the specific features of an aircraft track on the Patriot display. In particular, when task-specific features and prior information about a track provided an explanation for the most recently received, supposedly disconfirming information, the most recent information was reinterpreted as confirming, not disconfirming the explanation. The result was a primacy effect; prior information was weighted too heavily.

In contrast, when the track's features did not provide an explanation for the most recent disconfirming information, teams used one of two strategies. They either used (a) an anchoring and adjustment heuristic that weighted the most recent information too heavily, thereby resulting in a recency effect, or (b) an additive model that took into account all information, thereby resulting in no order effect. The type of effect, or the lack of one, depended on how the track looked on the Patriot display. Although they were not investigating the use of judgment heuristics, Boy (1995), Hammond (1988), Payne, Bettman, and Johnson (1993), and Sternberg and Wagner (1994) all discuss the context-dependent nature of human judgment.

In the second experiment, Adelman, Christian, Gualtieri, and Bresnick (1996) studied the effect of communication training and four group composition variables on Patriot teams' performance for two types of tasks: a task where there was conflicting

information about the one (or two) tracks on the Patriot display and the more routinely trained air defense task, where there are many tracks (e.g., more than ten) on the Patriot display, but no conflicting information about any track. It was predicted that communication training would significantly enhance communication quantity and quality and, in turn, team performance for both tasks. Although the training did sometimes improve team communication processes, it did not improve team performance.

The variable that had the biggest positive effect on communication quality and team performance was the number of hours a team had worked together. This effect was only found, however, for the type of task for which Patriot teams routinely train. It did not transfer to the less frequent and more cognitively stressing task where there is conflicting information about unknown aircraft. Our hypothesis is that it will take (1) longer training than time constraints permitted us to perform, and (2) training that emphasizes analysis of team members' judgment processes, and comparison with the processes deemed appropriate by training personnel, to improve performance in the "few track/cue conflict" task.

In the third experiment, which is described herein, we investigated the interaction between features of the computer display and how operators processed information. In an earlier study, Adelman, et al. (in press) tested the effectiveness of an additive display with Patriot operators. This display presented a pictorial representation of the relative importance weights being

used by the Patriot's additive algorithm. (The actual weights could not be presented because that information is classified for security reasons.)

We assumed that operators were trying to use the system's additive processing rule, but that they were overweighing the most recent piece of information, which resulted in an order effect. By showing them the system's additive rule, we were providing cognitive feedforward (Hammond, Stewart, Brehmer, & Steinmann, 1975); that is, telling them how they should process the information. Research reviewed in Balzer, Doherty, and O'Connor (1989) found that, in general, cognitive feedforward was an effective means of improving judgmental accuracy.

Although the display removed the order effect early in the tracks' history, it failed to do so late in the tracks' history. We hypothesized that this failure occurred because the additive display was inconsistent with how the operators' were processing information. In particular, we hypothesized that instead of using an additive model all of the time, operators were sometimes using explanation-based reasoning (Pennington and Hastie, 1993); that is, stories, or scripts of enemy or friendly aircraft behavior, to explain the patterns of information they had about the aircraft. This hypothesis was later supported in the first experiment performed on the current contract (Adelman et al., 1995a), and is also consistent with research by Cohen, Freeman, and Wolf (in press) and Klein (1993).

The purpose of the third experiment performed on the current

contract was twofold. First, we wanted to see whether we could develop a display format that would help people who were using explanation-based reasoning (e.g,, Patriot operators) be more additive in their thinking (e.g., like the Patriot system). We took the perspective that the additive processing rule used by the Patriot system was the correct way to make aircraft identification judgments. We wanted to see if we could get operators to use this rule, even though we knew they were sometimes using a different processing rule, simply by how we displayed how the system was reaching its judgments.

The second purpose of the study was to directly test the "display cognitive consistency hypothesis" proposed by Adelman, et al. (in press). This hypothesis states that the effectiveness of a display format for decision aiding systems, like Patriot, depends on the consistency between how the system displays its reasoning process and how the person is processing the information. If a person is using explanation-based reasoning and the system is using additive reasoning, then there is a basic inconsistency between the two.

We know from the Patriot study by Adelman, et al. (in press) that using an additive display designed by air defense experts will not result in more additive processing by Patriot operators, at least not late in the aircraft's flight path. The "display cognitive consistency hypothesis" predicts that the way to affect the operators' reasoning process is to display the system's reasoning process in a way that is consistent with how the

operator is processing the information. So, in order to get operators to use more additive processing, one needs to use "explanation-based reasoning displays;" that is, frame the presentation in terms of the operators' reasoning process.

The basic assumption for this hypothesis is that in situations where there is no performance feedback, the operators' reasoning process frames how they view the world. First, there is no outcome feedback providing the correct answer after the operators' decision for each aircraft; consequently, operators can not learn if they or the system is more accurate. Second, there is no cognitive feedback comparing how the system and person are making their judgments. Or as in the case with Patriot operators, they know but may disagree with the system's reasoning process for a particular track. Under these circumstances, there is no reason why the operators should adopt the system's judgment when it differs from their own. There is no way for operators to know that the system is more accurate. In addition, the rationale for the system's reasoning process is not presented in a manner that is consistent with the operator's reasoning process.

More broadly stated, the "display cognitive consistency hypothesis" states that performance depends on the similarity between how the system displays its reasoning and the operator's reasoning, even if the system's actual reasoning process is different than the operator's. In the case of the Patriot system, which uses an additive rule, this means that operators using explanation-based reasoning need an "explanation-based reasoning

display." Using an additive display, even though it is consistent with how the system is processing the information, will result in lower performance because it is inconsistent with how the operator is processing the information. The display needs to present the rationale for the system's additive processing in a manner that is consistent with how operators are processing information. Similarly, operators using an additive rule need an "additive display." Using an "explanation-based reasoning display" was predicted to result in lower performance, even though the recommendation for both displays was the same.

The "display cognitive consistency hypothesis" suggests the existence of a reasoning x display interaction. This interaction is shown pictorially in Figure 1, again assuming that the system is using an additive processing rule. The highest degree of agreement with the system's recommendation is predicted to occur when everything is consistent: the operator and the system use an additive rule and there is an additive display. The next highest level of agreement is when there is consistency between the operator's processing rule (i.e., explanation-based reasoning) and the display (i.e., an explanation-based reasoning display). Inconsistencies between the operator and the display are predicted to result in lower and equivalent levels of agreement with the system's recommendation. In all four cases, however, it was predicted that agreement with the system's recommendations would be higher than that obtained with information displays that provided no assistance (i.e., no recommendations).

- - - - - - - - - - - - - - - - - -

Insert Figure 1 about here

- - - - - - - - - - - - - - - - - -

Our perspective is consistent with a cognitive engineering approach to system design; that is, "... the design and development of computer-based information systems consistent with what we know about how humans process information and make decisions" (Andriole & Adelman, 1995, p. 10). However, we knew of no research testing the "display cognitive consistency hypothesis" with decision aiding systems.

Decision research reviewed in Kleinmuntz and Schkade (1993) shows that the characteristics of informational displays can significantly affect the type of decision processes people use, but these studies were not with decision aiding systems like Patriot, which automatically generate recommendations for the operator. In addition, none of these studies addressed the use of explanation-based reasoning or, for that matter, any non-compensatory combination rule. Although there is research (e.g., Adelman, Cohen, Chinnis, Bresnick, and Laskey, 1993) showing that the interface to an expert system like Patriot can significantly affect operators' decision processes and performance with the system, this research does not address the "display cognitive consistency hypothesis."

There is some research providing indirect support for the hypothesis. Again, the broader theoretical position is that in situations with no feedback, people use themselves (in this case,

how they process information) as a reference point that frames their evaluation of new information (in this case, the system's recommendation and the reasoning process supporting it). This position is consistent with the judgment theory developed by Sherif and Hovland (1961), who demonstrated that people judged the acceptability of attitudinal messages by the messages' similarity to their current position, and that a message's acceptability could be enhanced by how it was presented. Similarly, research on the framing of decision problems (e.g., Adelman, Gualtieri, and Stanford, 1995b; Tversky & Kahneman, 1981) indicates that how a person frames a decision problem significantly affects their decision processes including, at times, their willingness to deal with contradictory opinions (Russo and Schoemaker, 1989).

A cognitive cost-benefit perspective also supports the "display cognitive consistency hypothesis." Beach and Mitchell (1978) and Payne, Bettman, and Johnson (1993) have shown that people try to select decision strategies, at least in part, on the basis of a cognitive cost-benefit analysis; that is, they try to select strategies that will give them the most accuracy given the effort required to use them in a particular setting. Kleinmuntz and Schkade (1993) have argued that information displays that reduce the amount of effort to use a particular strategy (e.g., additive processing) are more likely to facilitate the use of that strategy and, therefore, lead to greater accuracy if the strategy is best for that situation.

From a cognitive cost-benefit perspective, one could predict that "display cognitive consistency" should reduce the amount of effort required to adjust one's strategy. That is, it should be easier for operators using explanation-based reasoning to change to an additive strategy when using an explanation-based reasoning display because the display frames the rationale for the recommendation in a manner consistent with how operators expect to see it.

However, one could take an opposite position. That is, one could argue that the additive display would be easier to use because it (1) makes the necessary calculations, (2) displays the results pictorially, and (3) requires less reading than the explanation-based reasoning display. From this perspective, one would expect faster decision times, and in turn, higher accuracy (because of the reduced cognitive effort) with the additive display even if operators are using explanation-based reasoning.

More generally, cognitive cost-benefit analysis argues that there should be an inverse relationship between cognitive effort (measured in the mean amount of time required to make a decision) and decision accuracy (measured by the amount of agreement with the display recommendation). This hypothesis is appropriate in decision situations where time is limited, such as in the air defense task, because one could easily imagine that, if there was ample time, better displays might help one think more critically (and take longer) before making a decision. Therefore, we also examined whether there was an inverse relationship between

decision time and agreement.

METHOD

This part of the paper is divided into five sections. The first section describes the research design. The second describes the participants. The third section enumerates the independent variables in the experiment. The fourth section describes the procedures used. The fifth section presents the dependent variables, and reviews our predictions.

Design

The design for the experiment was a 2 (Type of Training) x 3 (Type of Display) x 2 (Type of Track) factorial design. The two levels for type of training were use of an additive rule and use of explanation-based reasoning. The three levels on the type of display (where display type refers to the type of cognitive assistance provided by the display) were additive, explanation-based, and no decision assistance. The two levels on types of tracks were those tracks used in previous research with Patriot operators and "new tracks" developed for this study.

The training and display variables were between subject variables; the track type was a within subject variable. In all cases, the additive and explanation-based reasoning displays gave the same recommendation for each aircraft track, which was the recommendation of the additive model.

Participants

We were unable to work directly with Patriot operators because of their limited availability, and prior commitments to

do the experiment reported in Adelman, et al. (1996). As an alternative, we trained college students to perform a simulated Army air defense task developed using SuperCard on a Macintosh computer. Ninety (90) undergraduate students participated in the experiment, with fifteen (15) being randomly assigned to each of the six cells defined by the 2 (Types of Training) x 3 (Types of Display). The students were from either an introductory psychology or business statistics course, and participated in the experiment for course credit for a maximum of two hours.

Independent Variables

Training. There were two training conditions: additive and explanation-based. In both conditions, participants were first taught the cues (i.e., information) suggesting whether an aircraft track was hostile or friendly. For example, jamming friendly radar is a "hostile cue" because the aircraft is probably trying to prevent friendly radar from acquiring it well enough to fire a missile at it. In contrast, flying in the Safe Passage Corridor (SPC) is a "friendly cue" because that is where friendly aircraft should fly when they are within missile range.

In the Additive Training condition, participants were taught how to weight the cues to reach a judgement about the aircraft. Table 1 shows the weights for the Additive Processing Rule that the participants in this condition were trained to use, and that the system used in all conditions to make its recommendations. These are fictitious weights, they are not the weights used by the Patriot system, which are classified for security reasons.

Participants were given a short test to make sure that they could successfully use the weights to identify aircraft tracks shown by the system. Participants then proceeded to one of the display conditions where the experimental data was collected.

- - - - - - - - - - - - - - - - - -

Insert Table 1 about here

- - - - - - - - - - - - - - - - - -

In the Explanation-Based Reasoning (EBR) Training condition, participants were trained in various explanations for accounting for conflicting information about an aircraft. Explanation training did <u>not</u> include the use of the weights shown in table 1. Instead, the participants were taught four patterns of aircraft behavior. Two patterns explained why a friendly aircraft might perform specific hostile cues. The two patterns were called "cutting the corner" to explain why a friendly aircraft might leave the safe passage corridor, and "accidental jamming" to explain why a friendly aircraft might jam the operator's air defense radar. These patterns of behavior are shown pictorially in Figure 2.

- - - - - - - - - - - - - - - - - -

Insert Figure 2 about here

- - - - - - - - - - - - - - - - - -

Two other patterns explained why a hostile aircraft might perform friendly cues. The two patterns were called "bombing run on asset" to explain why a hostile aircraft might stop jamming the air defense radar late in the tracks history, and "corridor

guessing" to explain why a hostile aircraft might be in the safe passage corridor. These patterns of behavior are shown pictorially in Figure 3. As in the case of Additive Training, the participants in the Explanation-Based Training condition had to pass a test before proceeding to one of the three display conditions where the experimental data was collected.

- - - - - - - - - - - - - - - - - - - -

Insert Figure 3 about here

- - - - - - - - - - - - - - - - - - - -

Displays. Three types of displays were used in the experiment. Figure 4 shows what the No Decision Assistance Display looked like for a track late in its flight history.

- - - - - - - - - - - - - - - - - - - -

Insert Figure 4 about here

- - - - - - - - - - - - - - - - - - - -

There are three windows in the No Decision Assistance Display, each of which will be described in turn. The large window shows the graphic representation of the aircraft track (also called a target). The participants are playing the role of air defense operators, located at the bottom of the window, who are protecting two assets as well as themselves. The aircraft first appears at the top of the display at the Fire Support Coordination Line (FSCL), which participants were told separates enemy and friendly ground forces. The check mark within the circle indicates that the aircraft gave a friendly response (called a Mode 1, Mode 3) to an automatic, electronic

Interrogation Friend or Foe (IFF) inquiry. In addition, the aircraft is flying parallel to, but outside the friendly safe passage corridor.

The second piece of information the operator receives is that the target (i.e., aircraft) is jamming the operator's air defense radar. This is represented by the lightning bolt symbol. The aircraft continues flying outside the corridor until it turns into the mouth of the corridor. The aircraft then stops jamming, which is indicated by the line through the jamming symbol and the circle around it. Then the aircraft leaves the corridor, appearing to go toward the assets.

At this point, the participants had to make a decision to either "engage the track" because they thought it was hostile, or "clear the track" because they thought it was friendly, by clicking on the desired button at the top, right-hand corner of the display. Note that the track moves down this window in a series of "frames" so that each piece of information is presented sequentially. The engage and clear buttons do not appear on the screen until the decision point has been reached.

The window at the lower, left-hand corner of the display presents the cue information verbally. The window at the right-hand side of the display, below the engage and clear buttons, reminds the operator of the importance rank order (not weights) of the hostile and friendly cues, which they learned previously. Finally, the lower, right-hand window was left blank for the No Decision Assistance Display. It will be used in the Additive

Display and the Explanation-Based Reasoning Display.

The Additive Display is shown in Figure 5 for the same track shown in Figure 4. The Additive Display contains the same information as the No Decision Assistance Display, plus three additional features. First, below the clear and engage buttons, it presents the table of relative weights that the system is using to identify the track (i.e., Table 1). The weights are always present to make clear how the system is, and how the operator should, combine information to reach a decision. Second, after the last piece of information, the display presents the system's recommendation. For example, the rectangular box in the lower portion of the large window indicates that the system is recommending that the track be cleared.

- - - - - - - - - - - - - - - - - - -

Insert Figure 5 about here

- - - - - - - - - - - - - - - - - - -

Third, the lower, right-hand portion of the display pictorially represents the track's weighted score after each piece of information. The higher the score, the more friendly the aircraft according to the system's additive model. Friendly scores (those that were positive) were color coded in green; hostile scores were color coded red. The middle horizontal line indicates that point where the total score was zero; that is, neither in favor of friend nor hostile.

Additively-trained participants were trained to use the rule the system used to break ties. The rule was "Clear" if the IFF

Mode Response indicated a friend, and "Engage" if it indicated a hostile. Consequently, the system recommends that the track be cleared. The different shading in the figure is a function of how the different colored boxes looked when printed out, and should not be considered here.

The final display type was the Explanation-Based Reasoning Display. Figure 6 shows the Explanation-Based Reasoning Display for the track previously shown in Figures 4 and 5. This display varied three characteristics of the Additive Display, to make them appropriate for Explanation-Based Reasoning. Instead of listing the relative importance weights, it lists the rank order of the cues, just as in the No Decision Assistance Display.

Second, the Explanation-Based Reasoning Display distinguished between the system's identification and recommendation. We thought this might help the operators understand why the system recommended clearing or engaging an "unknown aircraft," which is one with a zero total score using the additive weights, without actually mentioning the tie-breaker rule, which would require a discussion of the additive rule.

- - - - - - - - - - - - - - - - - - - -

Insert Figure 6 about here

- - - - - - - - - - - - - - - - - - - -

Third, the explanation for the system's recommendation is presented in the lower, right hand corner. For this particular track, the system listed the two friendly cues (IFF Mode 1,3 and Stopped Jamming) and tried to explain-away the hostile cue, which

was being outside the safe passage corridor, as being due to navigational problems. Beginning with the third piece of information, the system began giving an explanation any time a new piece of information contradicted the current recommendation. We did not begin the explanations until the third piece of information to emulate a real system, which would need to collect some information before reaching even a tentative conclusion.

Explanations for the Explanation-Based Reasoning Display were developed using a set of rules to ensure that the same explanations were given to two aircraft with the same flight paths, but different order sequences. Where the explanation was given in the flight path depended on the order in which the information was presented, and the system's recommendation at that time. We developed these rules in an attempt to ensure that any obtained order effects were not a function of the explanations used by the system.

Tracks. There were two sets of tracks in the experiment: the 16 tracks previously used in the study by Adelman et al. (1996) with Patriot operators, which we will call "old tracks;" and 16 tracks developed for the current study, which we will call "new tracks." The old tracks were defined in terms of four two-level variables. The first variable was whether the initial information about the track was a friendly or hostile cue. The second variable was whether the track came down the left-hand or right-hand side of the display. The third variable was the late order information sequence; that is, whether the information late in

the track's history confirmed and then disconfirmed (CD) the initial information or disconfirmed and then confirmed the initial information (DC)).

These three, two-level variables (Initial Information, Side, and Late Order Sequence) were crossed with each other to create 8 tracks. The fourth "old tracks" variable was defined by different layouts of the friendly safe passage corridors. Each of the 8 tracks were represented in each layout, making a total of 16 old tracks. The same tracks looked different in each layout because of the relationship between the track's flight path and the configuration of the safe passage corridor.

The new tracks were developed to look different than the old tracks, and to increase the number of tracks considered by the participants. The new tracks used the same layouts as the old tracks, with eight tracks in each layout. The new tracks also manipulated the "information order sequence" to create four pairs of tracks for each layout, but did not restrict the order manipulation to late in the tracks flight path.

The new tracks did not manipulate the initial information. The initial information was a friendly cue for 10 of the 16 new tracks. The initial information for the other 6 tracks included both a friendly and hostile cue such that the initial score was 0 using the additive model.

The new tracks did not manipulate the side of the display either. Ten tracks came down the left-hand side, two came down the middle, and four came down the right-hand side of the

display. Since we wanted the new tracks to look significantly different from the old tracks, they tended to deviate more from the safe passage corridor than the old tracks. In addition, the new tracks used some cues not found in the old tracks. In general, the new tracks looked more hostile than the old tracks, even though some of them were clearly friends according to the additive model.

Procedures

The participants were first trained to understand the terms used in the air defense task, including the verbal definitions and pictorial representation of the cues. They were then tested on this material. The experimenter presented the correct answers for any questions answered incorrectly before proceeding further. The participants then used the No Decision Assistance Display to make engagement decisions for ten tracks. These tracks did not have conflicting information, and simply provided a means for letting the participants become familiar with performing the air defense task.

The participants then received either Additive or Explanation-Based Training, depending on their condition. They were tested on this material, and mistakes reviewed, before proceeding to a second computer training session. In the second session, the display gave assistance which emphasized the training condition. These training displays were different than those used to collect the data for the experiment.

Using the training displays, participants were given ten

"practice tracks" with conflicting information. The first three tracks were displayed at a much slower speed than the remaining seven to give the participants more time to make their decision. All participants had to get seven of the ten "practice tracks" correct to proceed into the display conditions. Only a few participants failed to proceed after two tries.

Participants were then trained in using the displays appropriate to their display condition. Participants using the Additive and Explanation-Based Reasoning Displays were told that the system was a prototype and, therefore, not always accurate when tracks had conflicting information. We did this for two reasons. First, to simulate real-world conditions. Previous research with Patriot operators indicates that some of them, whether correctly or incorrectly, think the Patriot algorithm sometimes arrives at the wrong decision when tracks have conflicting information. Second, and more importantly, we wanted to ensure that participants thought about the identification problem instead of simply doing whatever the system told them to do. The latter would not have been very interesting, nor representative of actual Patriot operators.

Dependent Variables and Predictions

There were two primary dependent measures in the experiment. The first measured the extent to which the participants agreed with the machine's recommendation, which was based on the additive model. Again, the purpose of the experiment was two-fold. The first purpose was to see whether we could develop a

display that would help people who were using explanation-based reasoning (e.g., Patriot operators) be more additive in their thinking (e.g., like the Patriot system.) Therefore, we took the perspective that the additive processing rule used by the Patriot system was the correct way to decide. This did not limit our ability to test the "display cognitive consistency hypothesis," which was the second purpose of the experiment. The predicted training x display interaction resulting from the hypothesis was shown previously in Figure 1.

The second dependent variable was the amount of time it took the participant to make the decision about the track. This dependent variable was used to as a surrogate for cognitive effort; for example, see Kleinmuntz and Schkade, 1993. This permitted us to test the general hypothesis that there is an inverse relationship between the amount of cognitive effort and the level of agreement with the display's recommendation. This is consistent with the "display cognitive consistency hypothesis." However, if one took the perspective that the Additive Display is faster to read than the Explanation-Based Reasoning Display, then one would predict longer decision times with the latter, regardless of its level of agreement.

RESULTS

A series of Analyses of Variance (ANOVAs) were completed to test whether the independent variables significantly affected the dependent variables. The results are presented for the two dependent variables, agreement with the recommendations of the

additive model and amount of time to make a decision, in turn. In addition, a number of post-hoc analyses are presented to clarify results where necessary.

Agreement

A 2 (types of training) x 3 (types of displays) x 2 (types of tracks) ANOVA was performed to test the "display cognitive consistency hypothesis." Training and display were between subject variables; tracks was a within subject variable.

There was a main effect for display: $F(2,84) = 4.39$, $MSe = 8.01$, $p = 0.015$. The mean agreement for the Additive Display was 76%. It was 75% for the Explanation-Based Reasoning (EBR) Display, and 67.5% for the No Decision Assistance (NDA) Display. These results indicate that, on the average, the Additive and EBR displays resulted in equivalent and higher performance than the No Decision Assistance Display. This was not particularly surprising since both types of displays gave decision assistance.

There was also a main effect for tracks: $F(1,84) = 4.33$, $MSe = 2.6$, $p = 0.041$. On the average, the agreement levels were higher for the "old tracks" (74.5%) than the "new tracks" (71.4%).

Contrary to our predictions, we did not obtain a significant training x display interaction. Nor did we find a main effect for training. However, the training x display x track interaction did approach significance: $F(2,84) = 2.97$, $MSe = 2.6$, $p = 0.057$. Figure 7 presents the mean values for the six training by display conditions for the "old tracks" and "new tracks" respectively.

- - - - - - - - - - - - - - - - - - - -

Insert Figure 7 about here

- - - - - - - - - - - - - - - - - - - -

There appears to be minimal difference between the mean values for the Additive and EBR displays for the "old tracks." Moreover, the slight differences that do exist appear to be counter to the hypothesis. In addition, there was little improvement over that obtained with the No Decision Assistance display.

In contrast, the pattern of results for the "new tracks" almost perfectly fit the predicted pattern shown in Figure 1. Consistent with the "display cognitive consistency hypothesis," the participants with additive training reached their highest level of agreement with the Additive Display. Participants with explanation-based training reached their highest level of agreement with the EBR Display. For both types of training, agreement was considerably lower, and in the predicted pattern, for the NDA Display. We were surprised, however, that the highest level of agreement was only 80%. It was reached with the Additive Display for the additively-trained participants.

In an effort to understand why the mean agreement values were so different for the two types of tracks, we examined the mean values for each of the six conditions for each of the 32 tracks used in the study. Not one of the sixteen "old tracks" fit the pattern of predicted results represented in Figure 1. More surprisingly, only four of the sixteen "new tracks" fit the

pattern, even though the mean values fit it.

What we found, instead, was large variation in the results for the six conditions as we moved from one track to another. This large variation indicated that the results were highly dependent on the specific set of circumstances (training, display, and track) operating in each case. We think we understand why this occurred. First, we present our post hoc hypotheses and second, the results to support them.

We maintain the position that one's training defines how one reacts to the recommendation and rationale for it, whether additive or explanation-based, but that "display cognitive consistency" must be considered on a track-by-track basis, not globally. In particular, one must consider three perspectives: (1) the type of decision that operators would make based solely on their training; (2) whether the system's recommendation and displayed rationale are consistent with that training; and (3) whether the pictorial representation of the aircraft (and any other information not considered in the system's recommendation and displayed rationale), is consistent with the training.

If everything is consistent with the operator's training, then one will obtain high levels of agreement with the system's recommendation. However, if the training, display (i.e., recommendation and rationale), or track picture are inconsistent, then two possible outcomes will occur. The first is that the system's recommendation will be discounted, and the operator will go with the decision that would be reached by training and/or the

effect of the track's representation. The second possibility is that the opposite will occur; that is, the recommendation will be accepted if the system can help the operator explain-away the contradictory evidence based on training.

We now present results supporting the situation-specific (i.e, track-by-track), "display cognitive consistency hypothesis," and demonstrating the large variation in the results obtained for specific tracks. We begin with Figure 8, which presents the results for one of the tracks (Track 71) in layout 2. Track 71 was the track shown in Figures 4, 5, and 6 for the No Decision Assistance Display, the Additive Display, and the Explanation-Based Reasoning Display, respectively.

- - - - - - - - - - - - - - - - - -

Insert Figure 8 about here

- - - - - - - - - - - - - - - - - -

The additive model indicates that Track 71 is a friend, and both the Additive and EBR Displays recommend that it be "cleared." The results obtained for this track closely resemble our predictions, with the exception that mean agreement for the "EBR Training + Additive Display" condition was lower than predicted. In fact, it was only 0.40. This is considerably lower than the mean agreement level of 0.867 achieved with "Additive Training + Additive Display." Why did this occur? Our answer is that the Additive Display's recommendation and rationale failed to address how the participants interpreted the information about Track 71 based on their training.

We will first describe the inferred decision process of an operator trained in explanation-based reasoning (EBR), and then show how the Additive Display for Track 71 (see Figure 5) is inconsistent with it. Specifically, based on EBR training, this aircraft looks like it's on a bombing run. Just as in the "bombing run pattern" shown in Figure 3, Track 71 has (1) stopped jamming, and (2) left the safe passage corridor going directly at the assets, late in its flight path. In addition, the aircraft appears to be "corridor guessing."

Figure 3 also shows how a hostile aircraft might look on the screen if it was "corridor guessing." In Figure 3, the aircraft is zigzagging around in the safe passage corridor. The key idea is that the hostile pilot doesn't quite know where the corridor is. The same assumption can be made for Track 71. Instead of zigzagging, however, this "presumed hostile" aircraft is going parallel to the corridor before it turns to attack the assets. In short, based on EBR training, Track 71 looks more like a hostile than a friendly aircraft.

The Additive Display does nothing to dissuade the EBR trained operator from this conclusion. The total score is 0, which means that the system used the tiebreaker (IFF Mode 1,3, which is the weakest friendly cue) to reach its "clear" recommendation. The aircraft's two other friendly cues, momentarily being in the corridor and stopping its jamming, can be explained-away as part of the bombing run by the EBR trained operator. Since the operator was told that the system was a

prototype and, therefore, not always correct, it's not too surprising in hindsight that the EBR trained operator did not agree with the recommendation.

In contrast, for the additively-trained operator, there is no inconsistency between training and the recommendation and rationale provided by the Additive Display. Both the operator and system are adding and subtracting the numbers (i.e., relative weights) for the aircraft. Both reach a total score of zero; in fact, the additive display helps the operators see how this total was reached and makes sure their arithmetic is correct. Moreover, the system's use of the tiebreaker is consistent with the additively-trained operators' training.

As predicted, EBR-trained operators performed considerably better for Track 71 when using the EBR Display (mean agreement = 0.733). However, for the EBR trained operators, the EBR Display fails to account for the "bombing run" explanation for the aircraft's flight path. And the aircraft does look like it's going right toward two assets. As a result, four of the fifteen EBR-trained operators engaged the aircraft. Global "display cognitive consistency" was not quite enough for perfect agreement; the display's specific explanations needed to account for alternate hypotheses based on training.

The need for situation-specific, display cognitive consistency is again illustrated in the next example. In particular, Figure 9 presents the results for Track 73, which the additive rule identifies as a friend using the tiebreaker of IFF

Mode 1,3. The mean agreement levels with the No Decision Assistance (NDA) Display were extremely low (0.333) for both types of training. Consistent with the "display cognitive consistency hypothesis," agreement was highest for the additively-trained operators with the Additive Display. However, contrary to our predictions, agreement for EBR-trained operators was also highest with the Additive Display. In fact, mean agreement for EBR-trained operators using the EBR Display was as low as that obtained with the NDA Display; that is, 0.333.

- - - - - - - - - - - - - - - - - -

Insert Figure 9 about here

- - - - - - - - - - - - - - - - - -

These results can be understood by examining the specific set of circumstances (training, display, and track) operating for this particular situation; that is, by taking a situation-specific focus. Figure 10 shows the EBR Display after the last piece of information for Track 73. Based on EBR training, this track looks like a hostile aircraft. In particular, it looks like it is corridor guessing prior to making a bombing run on the assets.

- - - - - - - - - - - - - - - - - -

Insert Figure 10 about here

- - - - - - - - - - - - - - - - - -

It's important to note that the direction pointer is directed toward the assets, not the middle of the corridor, even though the track is physically in the corridor. In addition, the

last piece of information is that the IFF was not operative. Although this is a neutral cue, it seems hostile as the last piece of information about the track, particularly if one thinks the aircraft is on a bombing run.

The EBR Display had no effect for EBR trained operators. Apparently, the explanation "aircraft may be damaged" had minimal if any effect. In addition, each of the friendly cues listed to support the recommendation could be explained-away. The IFF Mode 1,3 Response changed to Not Operative; stopped jamming is part of the bombing run; and the direction pointer suggests that the aircraft will no longer be in the corridor soon.

In contrast, the Additive Display was much more effective with EBR trained operators. Figure 11 shows the Additive Display for the last piece of information for Track 73. What is particularly interesting to note is the long positive slope for information in the second half of the track's flight path. This means that the track has been performing a number of friendly acts lately. In addition, the display clearly indicates that the IFF Not Operative is a neutral cue, not a hostile one.

- - - - - - - - - - - - - - - - - -

Insert Figure 11 about here

- - - - - - - - - - - - - - - - - -

The "friendly" activity late in the track's history may have caused the explanation-trained operators to see if they could explain-away the hostile activity early in the track's history. As it turns out, the track's jamming and sudden deviation from

the corridor, followed by it's "stop jamming" and subsequent return to the corridor, resembled a pattern we taught the participants; specifically, where a friendly aircraft was trying to jam hostile radar and take evasive action prior to being engaged by the enemy. Although post hoc, the "evasive action" pattern helps explain why EBR-trained operators had higher levels of agreement than additively-trained operators with the Additive Display.

Additively-trained operators did not know the "evasive action" pattern; consequently, they could not use it to explain-away the early hostile behavior of the track. These operators thought the track was hostile too. Mean agreement with the No Decision Assistance Display was just as low for the additively trained as the EBR trained operators.

We present one last set of track-specific results to support our post hoc hypothesis that the obtained results can only be understood by examining the specific set of circumstances (training, display, and track) operating for a particular situation; that is, by taking a situation-specific instead of a global focus. In addition, this last set of track-specific results further illustrates the large variation in results obtained with the Additive and EBR displays.

Figure 12 shows the results for Track 62, which is also classified as a friendly aircraft. In sharp contrast to the results for Track 73, the highest mean agreement levels for Track 62, for both types of training, were achieved with the EBR

Display. In fact, all fifteen additively-trained operators using the EBR Display made the same engagement decision as the additive rule. All the other results for Track 62 are consistent with the global "display cognitive consistency" hypothesis, including the high agreement level for the additively-trained operators using the Additive Display. The question is: why was performance so high with the EBR Display for this track?

- - - - - - - - - - - - - - - - - -

Insert Figure 12 about here

- - - - - - - - - - - - - - - - - -

To answer this question, one must first note that the agreement level of additively-trained operators using the No Decision Assistance Display was high (mean agreement = 0.80). Figure 13 presents the Additive Display for this track after the last piece of information. To additively-trained operators, and the additive system, there is no question that this track is friendly. Its final score is +4, and except for leaving the corridor and jamming briefly during the first half of its flight path, Track 62 has not performed a hostile cue.

- - - - - - - - - - - - - - - - - -

Insert Figure 13 about here

- - - - - - - - - - - - - - - - - -

The EBR Display supports the additively trained operators' focus on Track 62's friendly cues, and explains away the two hostile cues. Regarding the former, the explanation given by the EBR Display after the last piece of information for Track 62 is,

"Track history suggests friend: (1) IFF Mode 1,3, (2) Stopped jamming, (3) Target in SPC." More importantly, after the fourth piece of information, the EBR Display provides an explanation for why a friendly aircraft may have left the safe passage corridor and jammed briefly by saying, "Target may have detected hostile radar and performed evasive maneuver to avoid being shot down." In short, for this specific track, the EBR Display not only supported, but bolstered the decision process of additively-trained operators.

Decision Time

The second dependent variable considered in the experiment was the amount of time an operator took to make a decision; that is, click on the clear or engage button after they appeared on the display. As with the agreement dependent variable, a 2 (types of training) x 3 (types of display) x 2 (types of tracks) Analysis of Variance (ANOVA) was performed. Training and display were between-subject variables; tracks was within-subject.

There was a significant main effect for display: $F(2,84) = 5.67$, $MSe = 44.0$, $p = 0.005$. The mean decision time for the Additive Display was 5.12 seconds. It was 6.86 seconds for the No Decision Assistance Display, and 9.18 seconds for the EBR Display. These results suggest that, on the average, there was a minimal relationship between global agreement and decision time. The Additive Display had the fastest decision time, but did not achieve a significantly higher mean agreement level than the EBR Display. In fact, the mean decision time with the EBR Display was

79% slower than with the Additive Display. It was even 34% slower than the decision time with the No Decision Assistance Display, even though its mean agreement level was considerably higher.

There was also a significant main effect for tracks: $F(1,84) = 7.3$, $MSe = 3.35$, $p = 0.008$. The decision time for the "new tracks" was faster than the decision time for the "old tracks:" 6.68 vs. 7.42 seconds, respectively. This is opposite of the results obtained for agreement, where higher mean agreement levels were obtained for the "old tracks," not the "new tracks."

Lastly, there was a significant training x display x track interaction: $F(2,84) = 3.2$, $MSe = 3.35$, $p = 0.046$. Figure 15 presents the mean values for the six display x training conditions for the "old tracks," and for the "new tracks," respectively. Figure 14 shows that the Additive Display results in faster decision times for both types of training, for both types of tracks. Then comes the No Decision Assistance Display and lastly, the EBR Display. This ordering, as represented by the parallel lines, portrays the Display main effect pictorially. The interaction is caused by surprisingly slow decision times for the EBR Display for EBR-trained operators for the "old tracks."

- - - - - - - - - - - - - - - - - - - -

Insert Figure 14 about here

- - - - - - - - - - - - - - - - - - - -

It could have been that the global analysis presented above obscured the fact that there was an inverse relationship between agreement and decision time for individuals. Consequently, we

calculated the correlation between agreement and decision time for each of our 90 participants for the 32 tracks. A correlation of - 0.30 is significant at $p = 0.05$ with $df = 30$. Table 2 shows the number of participants that had correlations lower than or equal to -0.30, between -0.30 and 0.0, and greater than or equal to 0.0 for each of the six training x display conditions.

- - - - - - - - - - - - - - - - - - -

Insert Table 2 about here

- - - - - - - - - - - - - - - - - - -

Only 23% of the participants had significant correlations at the $p \leq 0.05$ level. We did a number of chi squares, but no matter how we collapsed the data shown in Table 2, the chi squares were not significant at the $p = 0.05$. Therefore, we conclude that, in total, there was minimal support for the predicted inverse relationship between agreement and decision time.

DISCUSSION

Contrary to our prediction, we did not obtain support for the global "display cognitive consistency hypothesis." Instead, we found large variation in the agreement levels of differentially trained operators using different types of displays. Sometimes the mean agreement levels for a specific track resembled the predictions of the "display cognitive consistency hypothesis," with the highest level of agreement being achieved when the type of decision display (Additive or EBR) was consistent with the type of training (additive or explanation-based.) More often than not, however, track-specific

results did not fit any predetermined global pattern.

Careful examination of the track-specific results, and of the consistency of a display's recommendation and rationale with an operator's training, suggest that one must take a situation-specific focus to "display cognitive consistency," not a global one. That is, one must consider the specific set of circumstances (training, display, and track) to understand the results.

This distinction is critical because there is a tendency to assume that people will use one predominant reasoning process, and that in order for performance to be high, one must engineer the display consistent with that process. The study by Adelman et al. (1995a) showed, however, that trained operators use more than one judgment process, and that situation-specific characteristics trigger when different judgment processes are used.

The current study indicates that, even when operators are trained to have only one predominate reasoning process, situation-specific circumstances might cause a system's displayed rationale for a decision recommendation to be ignored because in that particular case, it is inconsistent with the operator's training. Again, the broader theoretical position is that in situations with no feedback, people use themselves (in this case, how they process information) as a reference point that frames their evaluation of new information (in this case the system's recommendation and the reasoning process supporting it.)

To implement a situation-specific focus, one must examine three perspectives. First, one needs to consider the type of

decision that operators would make based solely on their training. That is, we still take the position that one's training (or more broadly, task experience) defines how operators initially frame their decision based on the available information. Second, one must consider whether the system's recommendation and displayed rationale for each specific case is consistent with the operator's training. And, third, one must consider whether other information not considered in the system's recommendation and displayed rationale (e.g., the pictorial representation of the aircraft on the display) is consistent with that training.

If everything is consistent with the operator's training, then one will obtain high levels of agreement with the system's recommendation. If they are inconsistent, then our results suggest that, on the average, one of two possible thing will happen. Either the system's recommendation will be discounted, and the operator will go with the decision that would be reached by training and/or by the effect of other information. Or the opposite will occur; that is, the system's displayed recommendation will be accepted if the displayed rationale can help the operator explain-away the contradictory evidence based on training.

Contrary to the predictions of a global "display cognitive consistency hypothesis," or a cognitive cost-benefit analysis, having displays that are consistent with one's training and experience does not result in faster decision times. Instead, the

results support the hypothesis that the Additive Display was simply faster to read and react to than the EBR Display, with the result being faster decision times with the former regardless of the operator's training, or level of agreement with the recommendations of the additive rule.

The Additive Display provided a cumulative point score evaluation after the presentation of each piece of information; consequently, the recommendation (and score) after the last piece of information could be considered quickly. The decision time with the NDA Display was somewhat slower because the operator was solely responsible for aggregating the information into a decision. Nevertheless, it was faster, on average, than having the recommendation and rationale from the EBR Display.

The EBR Display clearly took the most time to read and to consider the implications of the explanations presented for the recommendation. The particularly long decision times with the EBR Display for explanation trained operators suggests that it takes longer to compare the viability of competing explanations, one set from experience and one from the system, at least with the interface used in this study.

We conclude by noting that the operators were faced with an extremely difficult task. The tracks had conflicting information; there was no way to obtain additional information about the tracks; and there was no outcome or cognitive feedback. Consequently, operators were in a situation where there was no way to learn how well they, or the decision aid (called a

prototype), was performing.

Although such situations are difficult and infrequent, they can occur. For example, the experimental task represents an Army air defense situation where there is conflicting information about an incoming aircraft track and the Patriot battery is operating in autonomous mode because communications with headquarters have been disrupted. Tragically, the task can also represent a naval air defense situation; in particular, the situation facing the U.S.S. Vincennes when it shot-down an Iranian airliner in 1987.

Such real world situations represent difficult system design problems, for as Vicente, Christoffersen, and Pereklita (1995, p. 529) point out, "... they are characterized by events which are unfamiliar to operators and that have not been anticipated by designers." Although we have taken a cognitive engineering perspective in an effort to address such design problems, the results presented herein suggest that a global cognitive engineering perspective will not work. Instead, our post hoc hypothesis, which needs additional evaluation using controlled experiments, is that one must take a situation-specific focus.

A situation-specific focus is consistent with the results of decision research performed in naturalistic settings (e.g., see Adelman et al., in press; Cohen et al., in press; and Klein, 1993), and with process control research reported in Vicente et al., (1995) and Rasmussen and Vicente (1990). However, a situation-specific focus makes the system designer's task more

difficult. Although we have learned alot about how people process information and make decisions, substantially more research is needed before we can provide designers with reliable guidance on how to design displays for situations where the decision events are unanticipated by the operator and designer alike.

REFERENCES

Adelman, L., and Bresnick, T.A. (1992). Examining the effect of information sequence on Patriot air defense officers' judgments. Organizational Behavior and Human Decision Processes, 53, 204-228.

Adelman, L., Bresnick, T., Black, P., Freeman, F.F., & Sak, S.G. (in press). Research with Patriot air defense officers: Examinining information order effects. Human Factors.

Adelman, L., Bresnick, T., Christian, M., Gualtieri, J., & Minionis, D. (1995a). Demonstrating the effect of context on order effects. Fairfax, VA: George Mason University.

Adelman, L., Christian, M., Gualtieri, J., & Bresnick, T. (1996). Demonstrating the effects of communication training and team composition on the decision making of Patriot air defense teams. Fairfax, VA: George Mason University.

Adelman, L., Cohen, M.S., Bresnick, T.A., Chinnis, J.O. Jr., & Laskey, K.B. (1993). Real-time expert system interfaces, cognitive processes, and task performance: An empirical assessment. Human Factors, 35, 243-261.

Adelman, L., Gualtieri, J., & Stanford, S. (1995b). Examining the effect of causal focus on the option generation process: An experiment using protocol analysis. Organizational Behavior and Human Decision Processes, 61, 54-66.

Adelman, L., Tolcott, M. A., & Bresnick, T.A. (1993). Examining the effect of information order on expert judgement. Organizational Behavior and Human Decision Processes, 56, 348-369.

Andriole, S., & Adelman, L. (1995). Cognitive systems engineering for user-computer interface design, prototyping, and evaluation. Hillsdale, NJ: Lawrence Erlbaum.

Balzer, W.K., Doherty, M.E., & O'Connor, R. (1989). Effects of cognitive feedback on performance. Psychological Bulletin, 106, 410-433.

Beach, L.R., & Mitchell, T.R. (1978). A contingency model for the selection of decision strategies. Academy of Management Review, 3, 439-449.

Boy, G. (1995). "Human-like" system certification and evaluation. In J.M. Hoc, P.E. Cacciabue, & E. Hollnagel (Eds.), Experts and Technology: Cognition & Human-Computer Cooperation (pp. 243-254). Hillsdale, NJ: Lawrence Erlbaum.

Cohen, M.S., Freeman, J.T., & Wolf, S. (in press). Meta-recognition in time-stressed decision making: Recognizing, critiquing, and correcting. Human Factors.

Hammond, K.R. (1988). Judgment and decision making in dynamic tasks. Information and Decision Technologies, 14, 3-14.

Hammond, K.R., Stewart, T.R., Brehmer, B., Steinmann, D.O. (1975). Social judgment theory. In M.F. Kaplan & S. Schwartz (Eds.), Human judgment and decision processes. New York: Academic Press.

Heath, L., Tindale, R.S., Edwards, J., Posavac, E.J., Bryant, F.B., Henderson-King, E., Suarez-Balcazar, Y., & Myers, J., (Eds.). (1993). Applications of heuristics and biases to social issues. New York: Plenum Press.

Klein, G.A. (1993). A Recognition-Primed Decision (RPD) model of rapid decision making. In G.A. Klein, J. Orasanu, R. Calderwood, & C.E. Zsambok (Eds.), Decision making in action: Models and methods (pp. 138-147). Norwood, NJ: Ablex.

Kleinmuntz, D.N., & Schkade, D.A. (1993). Information displays and decision processes. Psychological Science, 4, 221-227.

Payne, J.W., Bettman, J.R., & Johnson, E.J. (1993). The adaptive decision maker. New York: Cambridge University Press.

Pennington, N., & Hastie, R. (1993). A theory of explanation-based decision making. In G.A. Klein, J. Orasanu, R. Calderwood, and C.E. Zsambok (eds.), Decision making in action: Models and methods (pp. 188-201). Norwood, NJ: Ablex.

Rasmussen, J., & Vicente, K.J. (1990). Ecological interfaces: A technological imperative in high-tech systems. International Journal of Human-Computer Interaction, 2, 93-110.

Russo, J.E., & Shoemaker, P.J.H. (1989). Decision traps: The ten barriers to brillian decision-making & how to overcome them. New York: Doubleday.

Sherif, M., & Hovland, C.I. (1961). Social judgment: Assimilation and contrast effects in communication and attitude change. New Haven: Yale University.

Sternberg, R.J., & Wagner, R.K., (Eds.). (1994). Mind in Context. New York: Cambridge University Press.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. Science, 211, 453-458.

Vicente, K.J., Christoffersen, K., & Pereklita, A. (1995). Supporting operator problem solving through ecological interface design. IEEE Transactions on Systems, Man, and Cybernetics, 25, 529-544.

FOOTNOTES

[1] The research described herein was supported by contracts from the Research & Advanced Concepts Office of the U.S. Army Research Institute (ARI), Contract No. MDA903-92-K-0134, and the AASERT Program sponored by the U.S. Army Research Office (ARO), DAAH04-93-G-0286 (32239-RT-AAS), to George Mason University (GMU). We would like to acknowledge the effort of the following research assistants who contributed to this endeavor: Laura Bowling, Mary Anne Flood, Ken Oberg, and Lawrence Sklar. The views, opinions, and findings contained herein are those of the authors and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other official documentation.

[2] Reprint requests should be addressed to Leonard Adelman, Dept. of Operations Research and Engineering, School of Information Technology and Engineering, George Mason University, Fairfax, VA 22030-4444.

Table 1

Relative Importance Weights for Additive Training Condition and Additive Display

Hostile Cues	
Starts Jamming	-4
Leaves Safe Passage Corridor (SPC)	-3
Out of SPC at the Fire Support Coordination Line (FSCL)	-2
Pop-Up	-2
IFF No Mode	-1
Friendly Cues	
Stops Jamming	+4
Not Jamming at FSCL	+2
Enters SPC	+2
IFF Mode 4	+2
IFF Mode 1,3	+1
Neutral Cues	
IFF not operative	0
Enters defense zone in SPC	0

Table 2

Agreement by Decision Time Correlations Organized By the Six Training x Display Conditions

Type of Training	Type of Display	Number of Participants With $r \leq -0.30$	$-0.30 < r < 0.0$	$r \geq 0.0$
	Additive	1	8	6
Additive	EBR	5	8	2
	No Assist.	7	5	3
	Additive	1	9	5
EBR	EBR	4	6	5
	No Assist.	3	5	7
Total Number		21	41	28
Total Percentage		23%	46%	31%

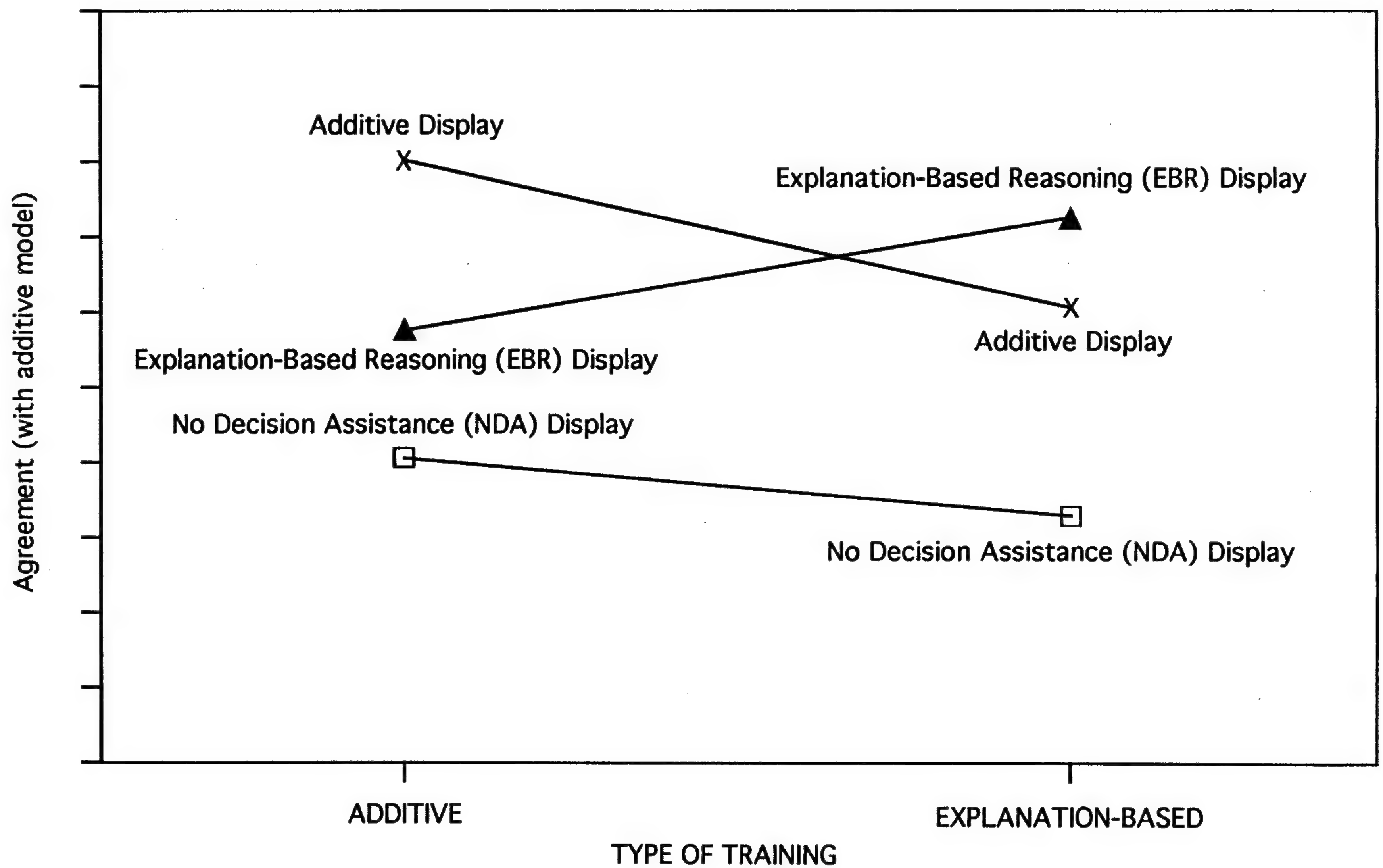
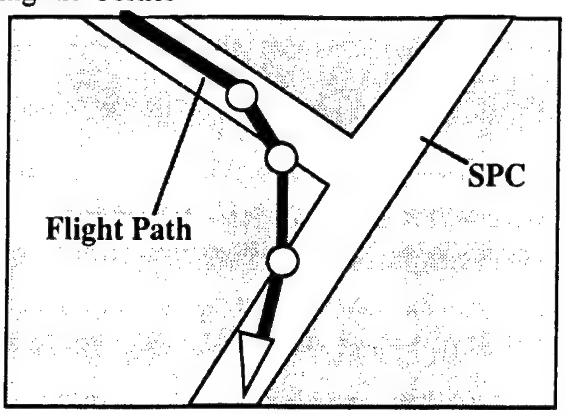


Figure 1. Predicted Training x Display interaction.

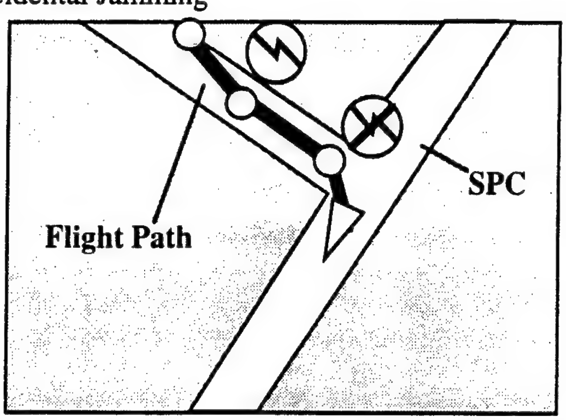
I. Cutting the Corner



1. Aircraft is in safe passage corridor (Neutral Cue)
2. Aircraft leaves SPC (Hostile Cue)
3. Aircraft returns to SPC (Friendly Cue)

Explanation: Pilot is either sloppy or in a hurry and is taking the turn too tight

II. Accidental Jamming

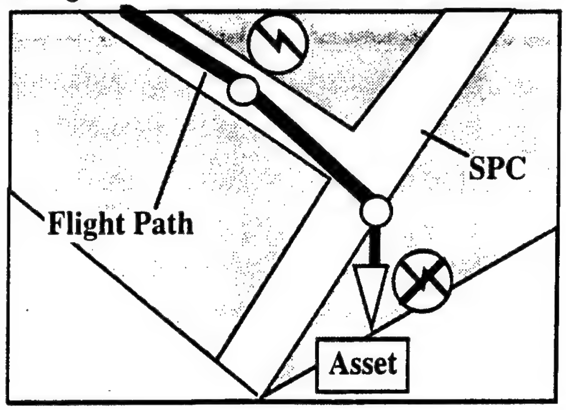


1. Aircraft enters safe passage corridor (SPC) (Friendly Cue)
2. Aircraft starts jamming (Hostile Cue)
3. Aircraft stops jamming (Friendly Cue)
4. Aircraft continues in SPC (Neutral Cue)

Explanation: Pilot accidently turned on jammers or thought that a hostile radar locked on to the aircraft

Figure 2. Two patterns used in Explanation-Based Reasoning (EBR) Training to explain why a friendly aircraft might perform specific hostile cues.

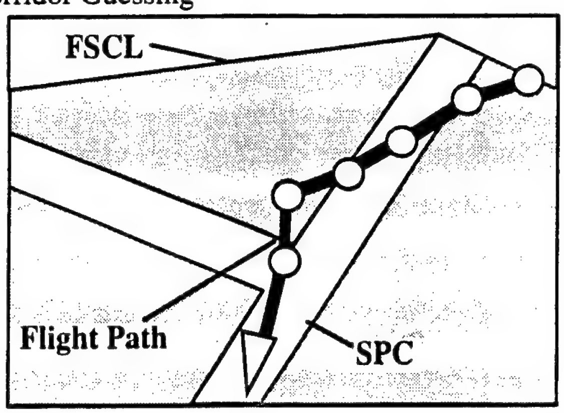
III. Bombing Run on Asset



1. Aircraft starts jamming (Hostile Cue)
2. Aircraft leaves SPC (Hostile Cue)
3. Aircraft stops jamming (Friendly Cue)

Explanation: Aircraft is hostile and has stopped jamming so that it can use its weapons to get a radar lock on the asset and attack it.

IV. Corridor Guessing



1. Aircraft crosses FSCL outside safe passage corridor (Hostile Cue)
2. Aircraft enters SPC (Friendly Cue)
3. Aircraft continues flying in SPC (Neutral Cue)
4. Aircraft leaves SPC (Hostile Cue)
5. Aircraft turns back toward SPC (Neutral Cue)
6. Aircraft returns to SPC (Friendly Cue)

Explanation: Aircraft is hostile and is trying to copy the flight path of other aircraft that it has seen using the safe passage corridor.

Figure 3. Two patterns used in Explanation-Based Reasoning (EBR) Training to explain why a hostile aircraft might perform specific friendly cues.

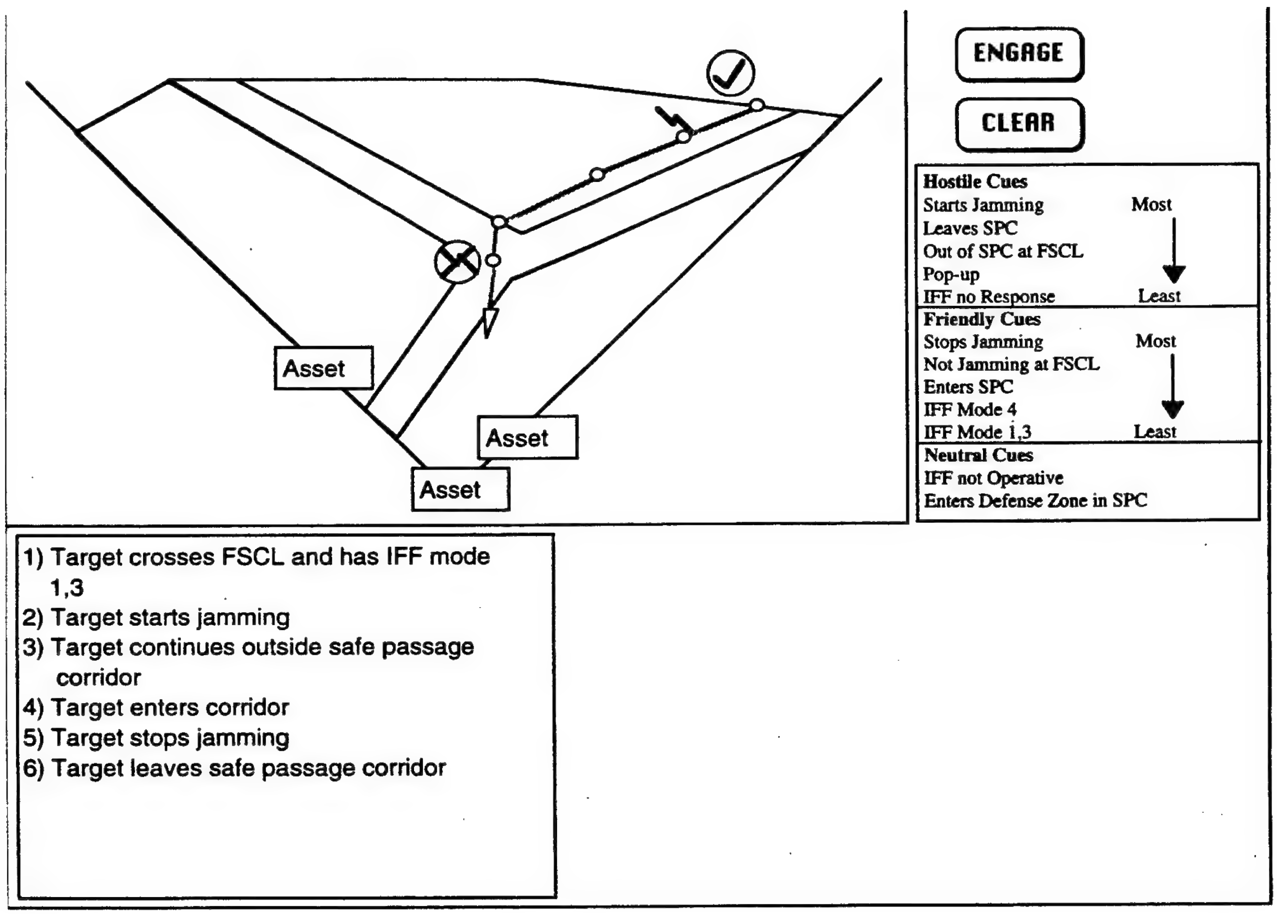


Figure 4. How the No Decision Assistance (NDA) Display looked after the last piece of information for Track 71.

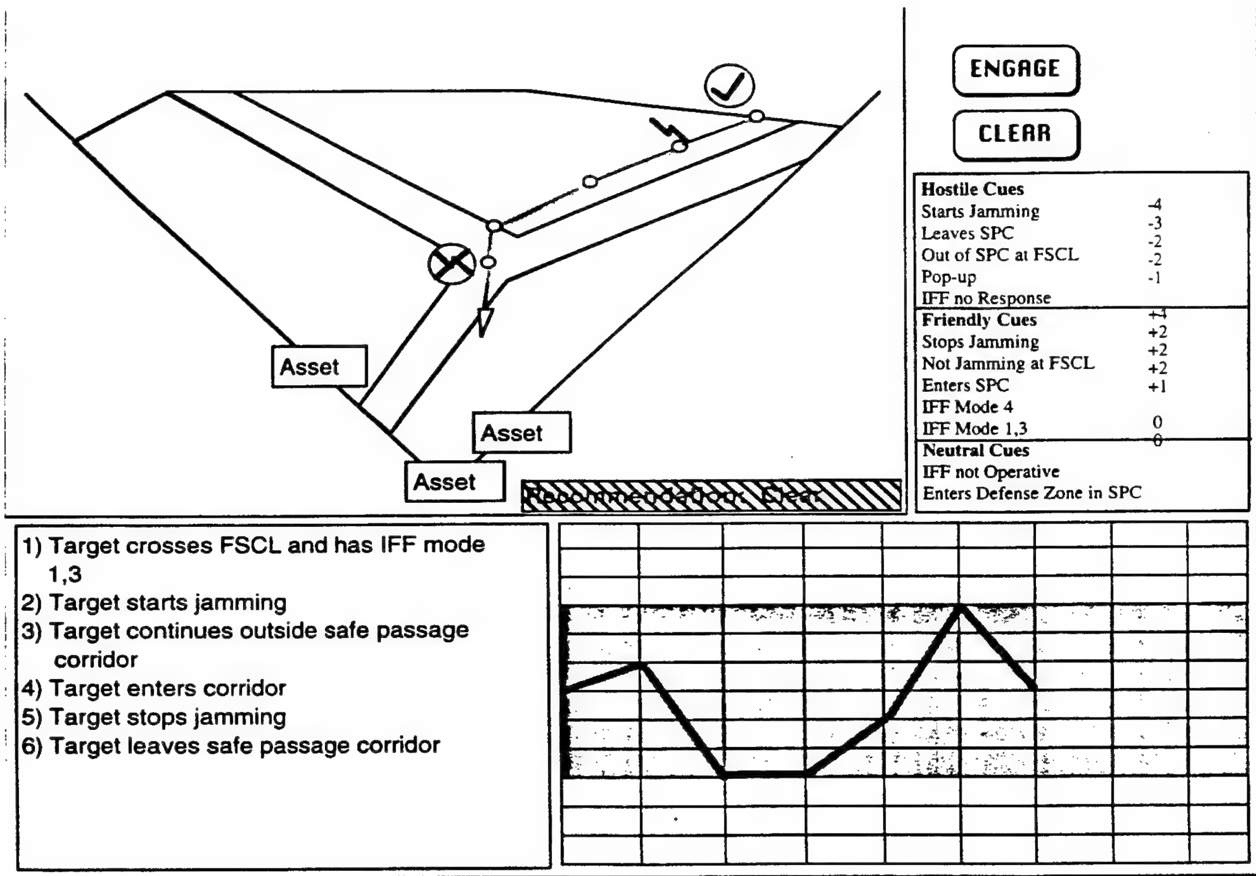


Figure 5. How the Additive Display looked after the last piece of information for Track 71.

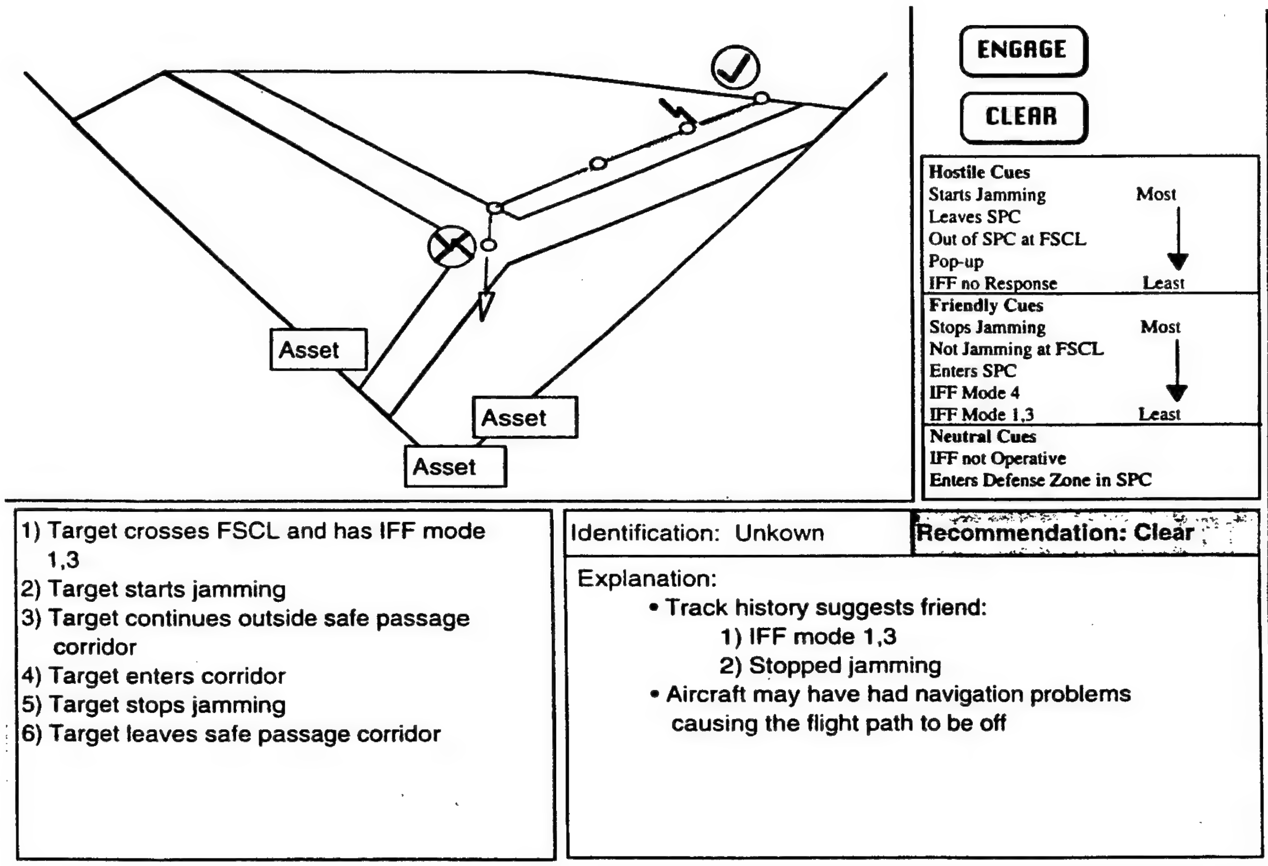


Figure 6. How the Explanation-Based Reasoning (EBR) Display looked after the last piece of information for Track 71.

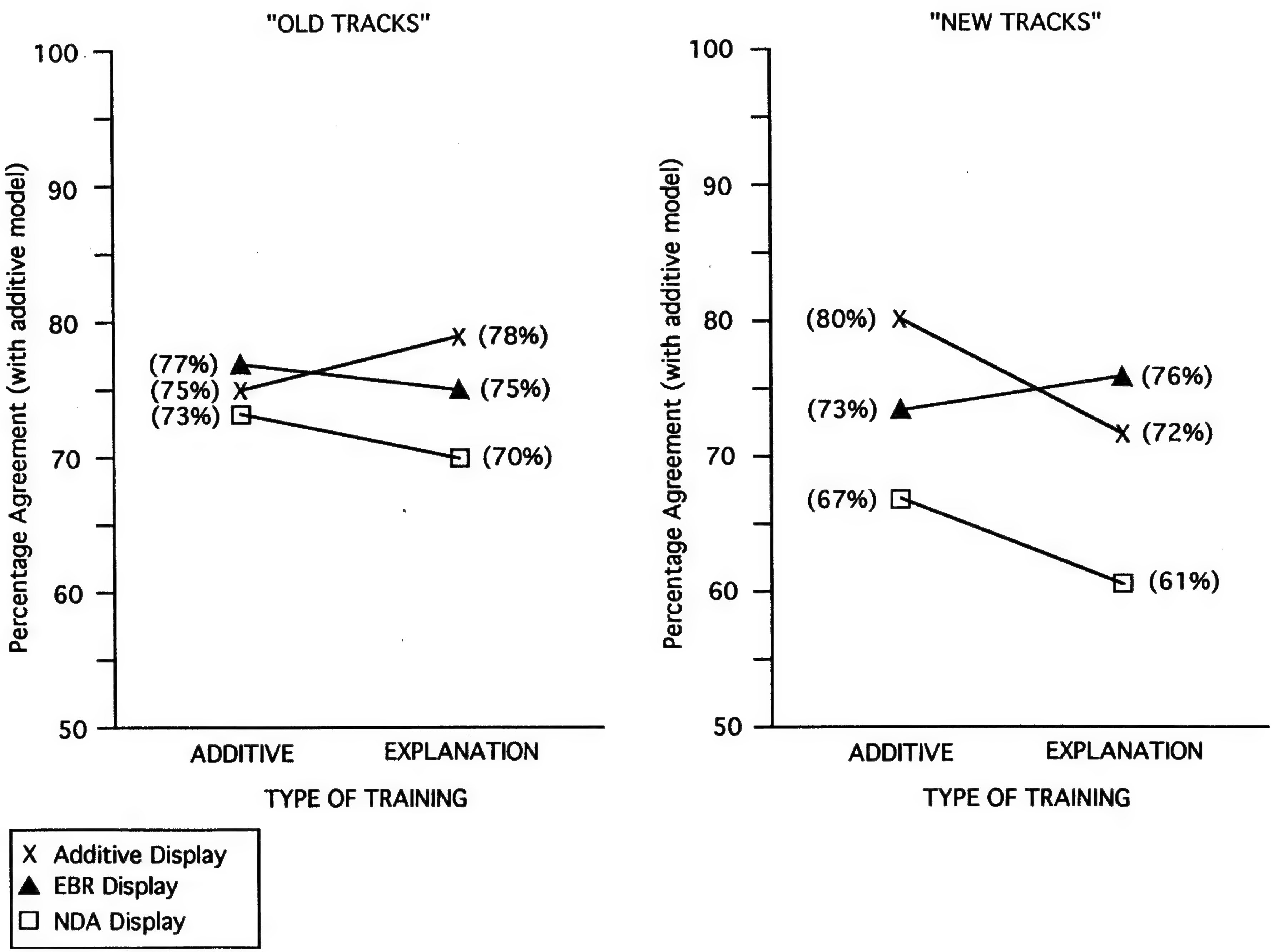


Figure 7. Mean agreement for six conditions, for each of the two types of tracks.

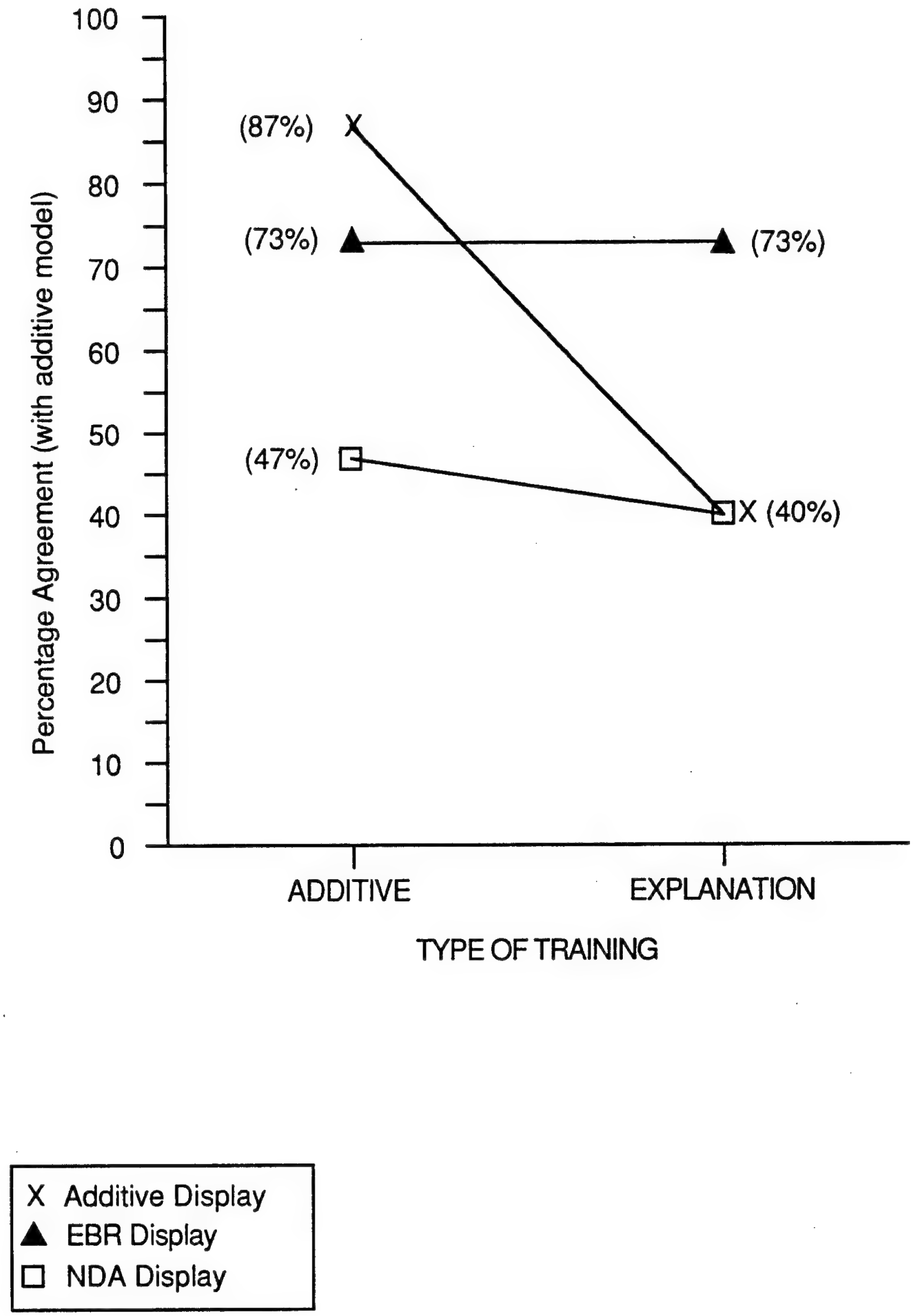


Figure 8. Results for Track 71.

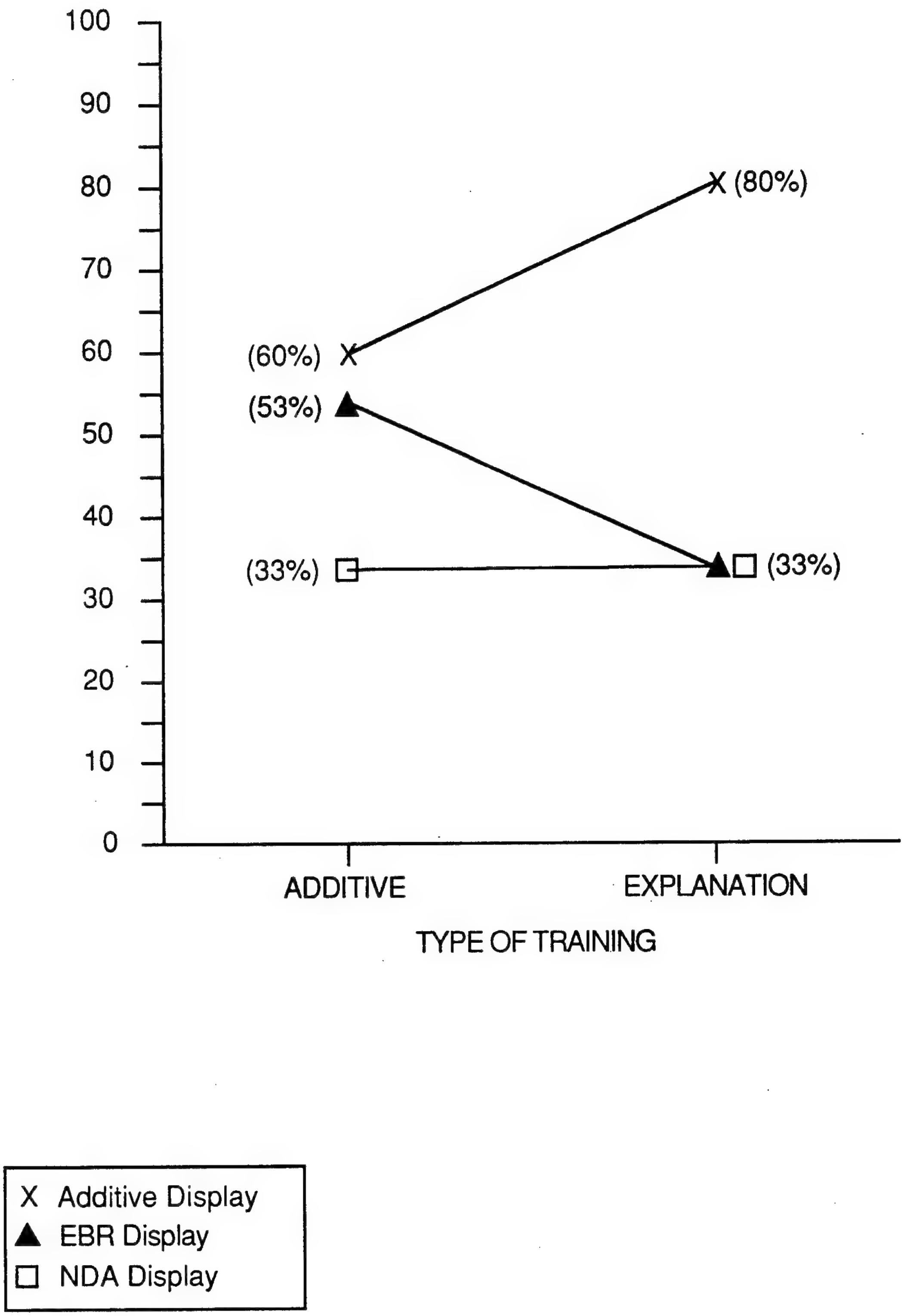


Figure 9. Results for Track 73.

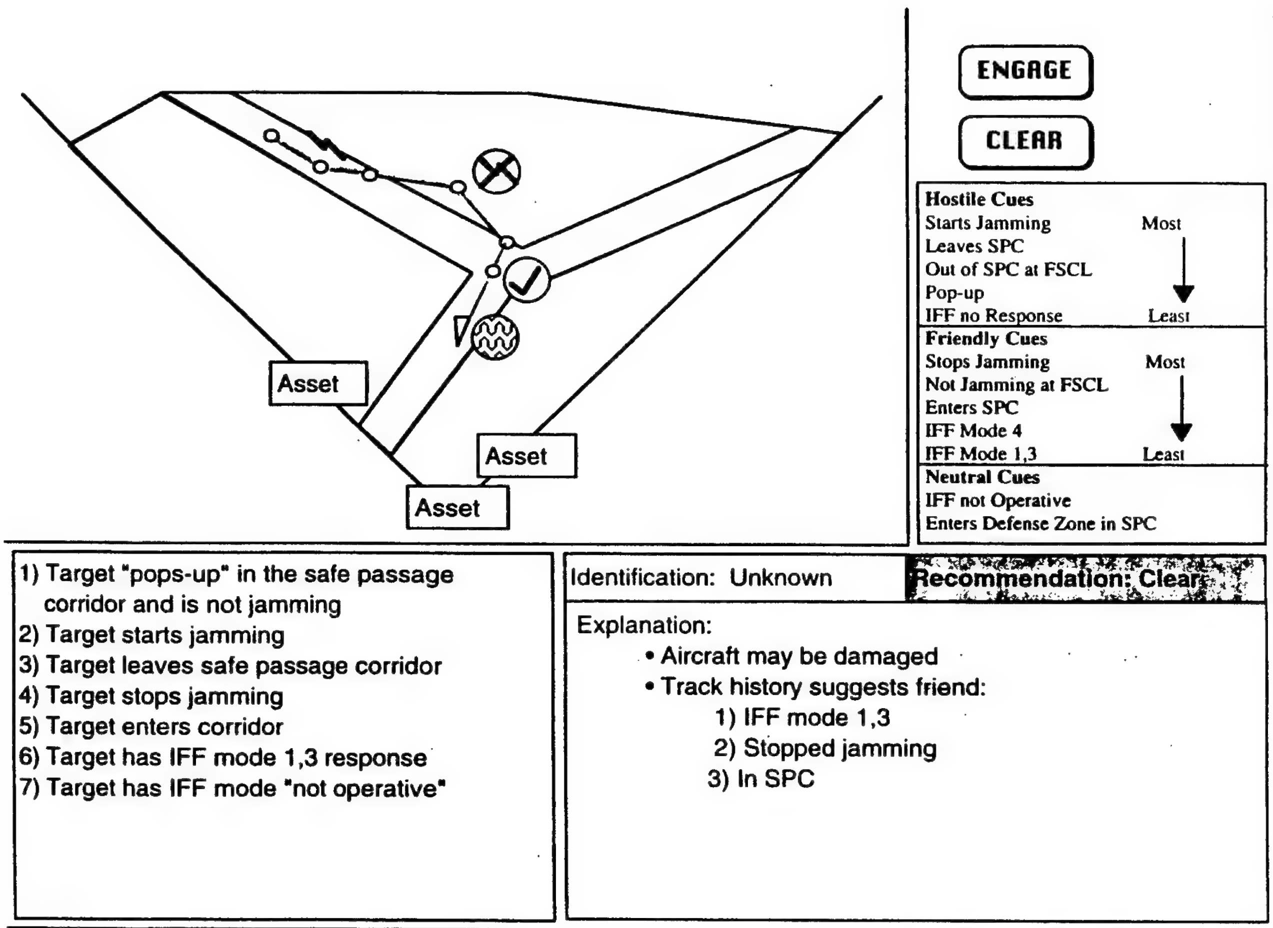


Figure 10. How the EBR Display looked after the last piece of information for Track 73.

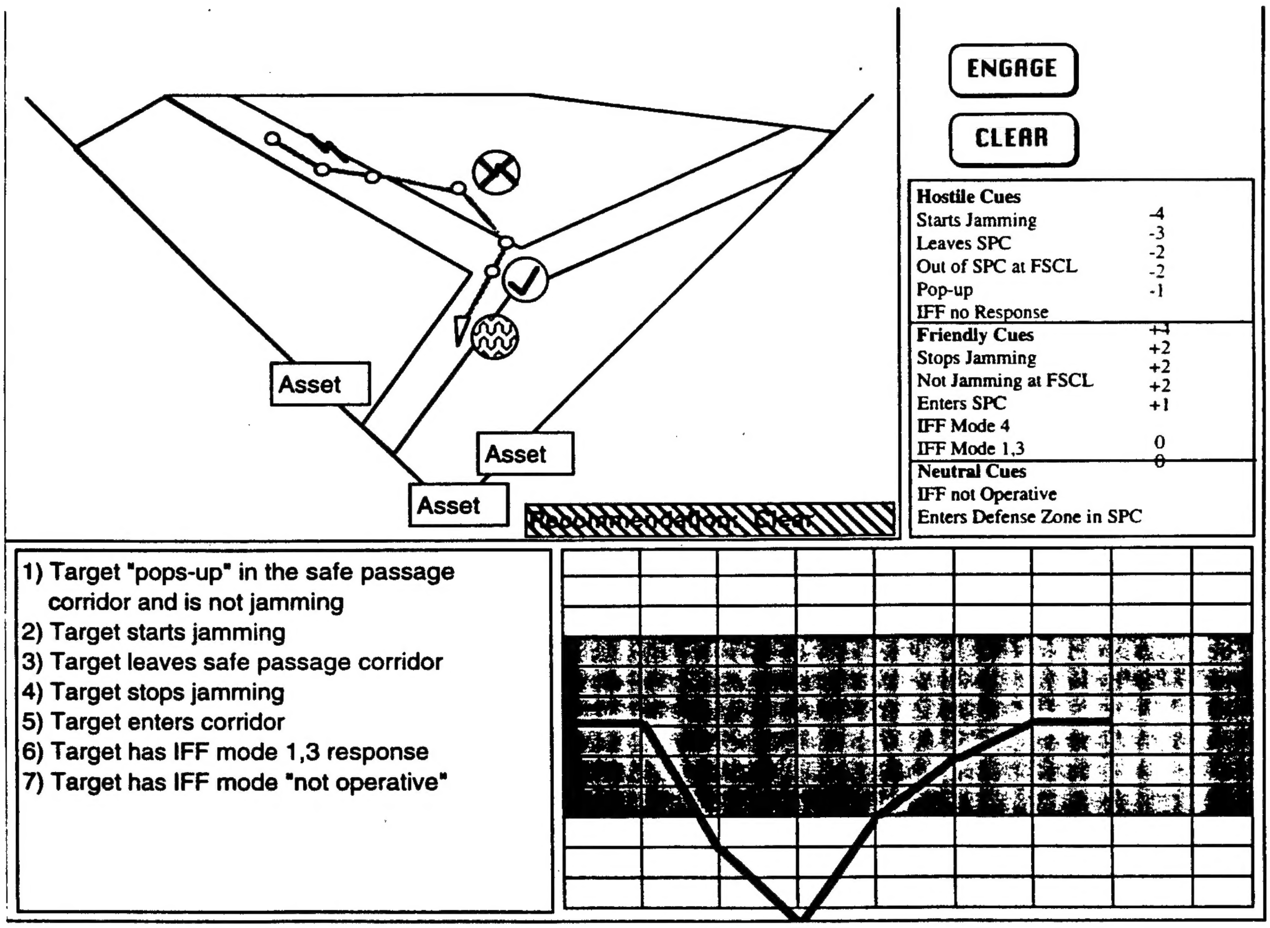


Figure 11. How the Additive Display looked after the last piece of information for Track 73.

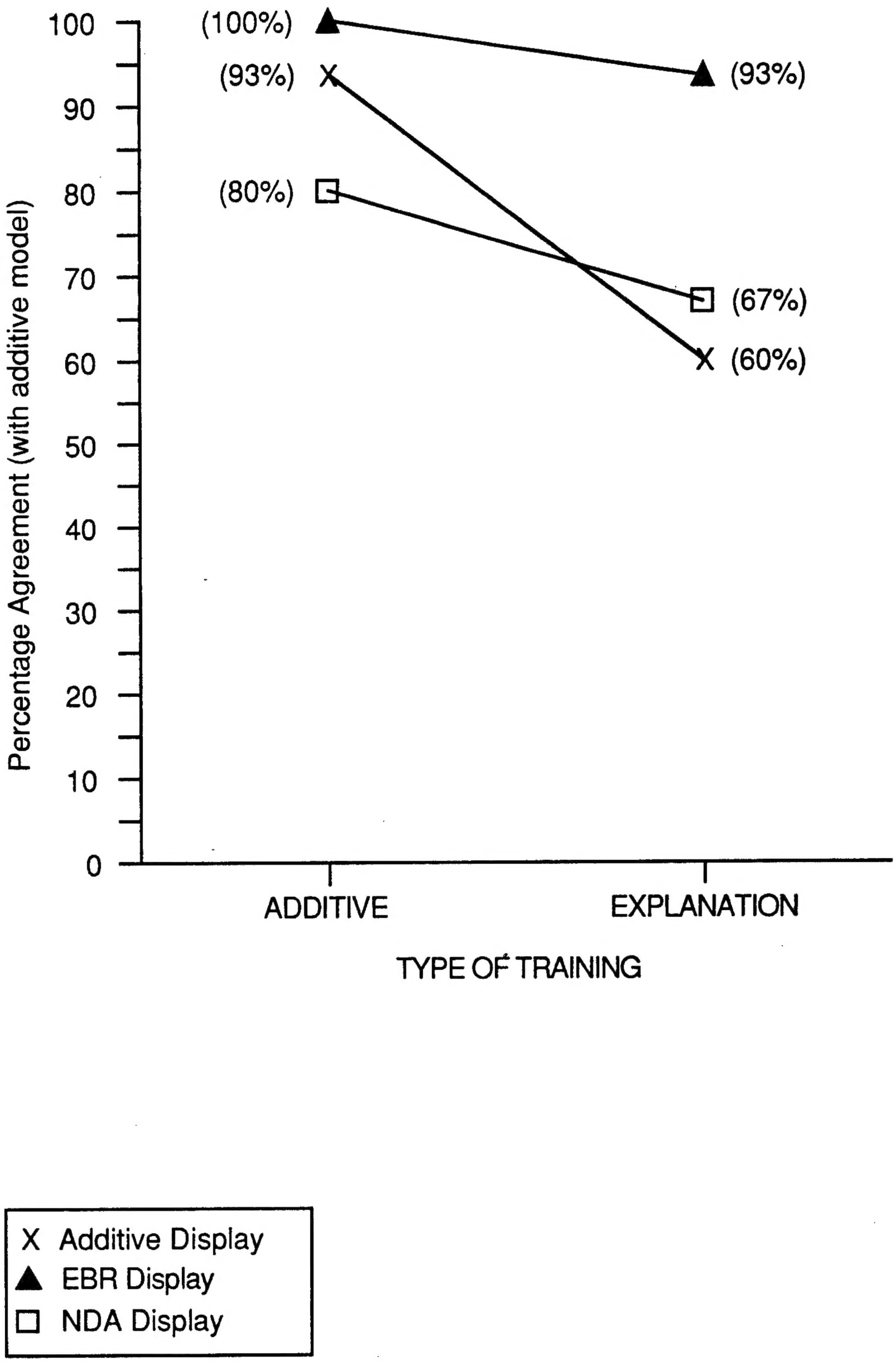


Figure 12. Results for Track 62.

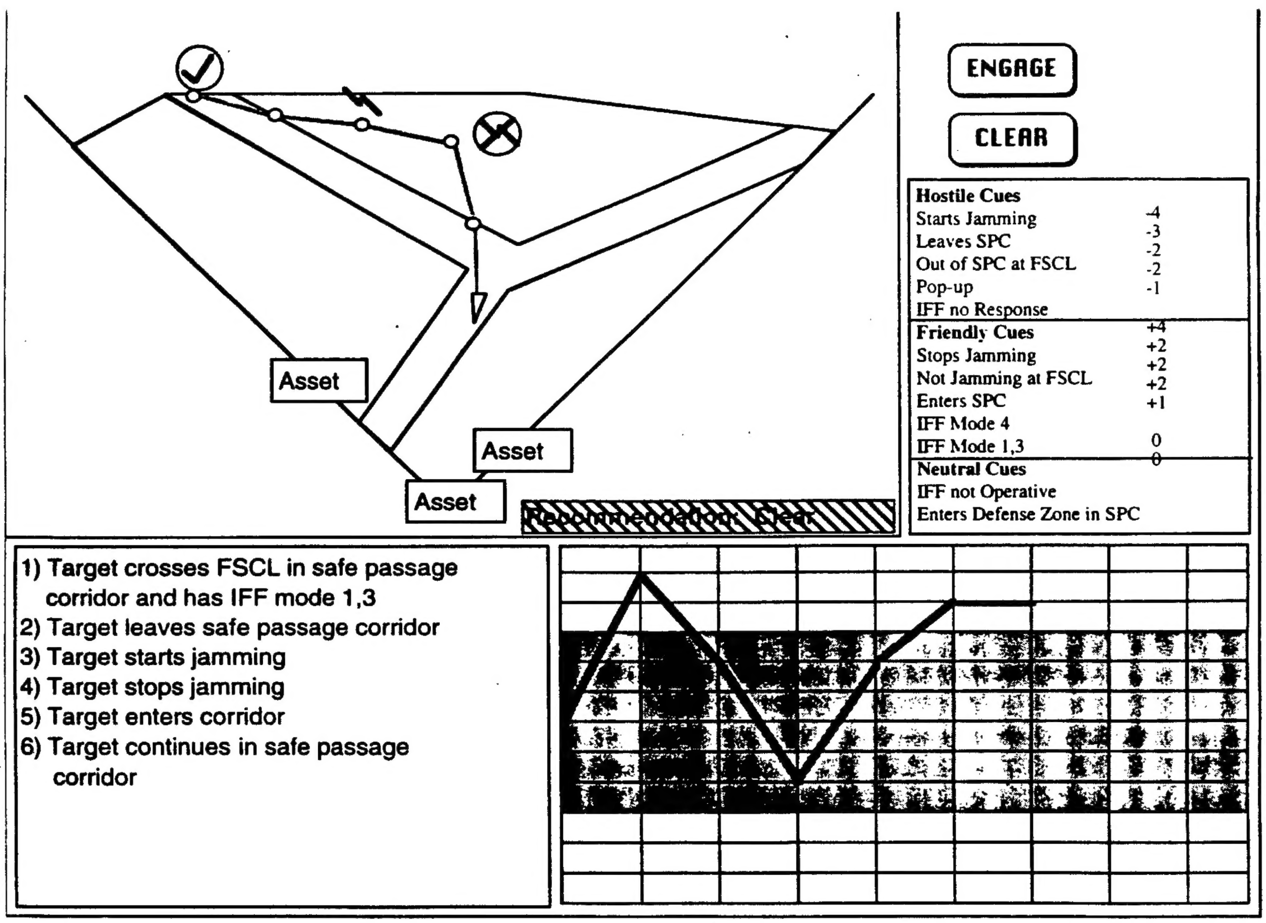


Figure 13. How the Additive Display looked after the last piece of information for Track 62.

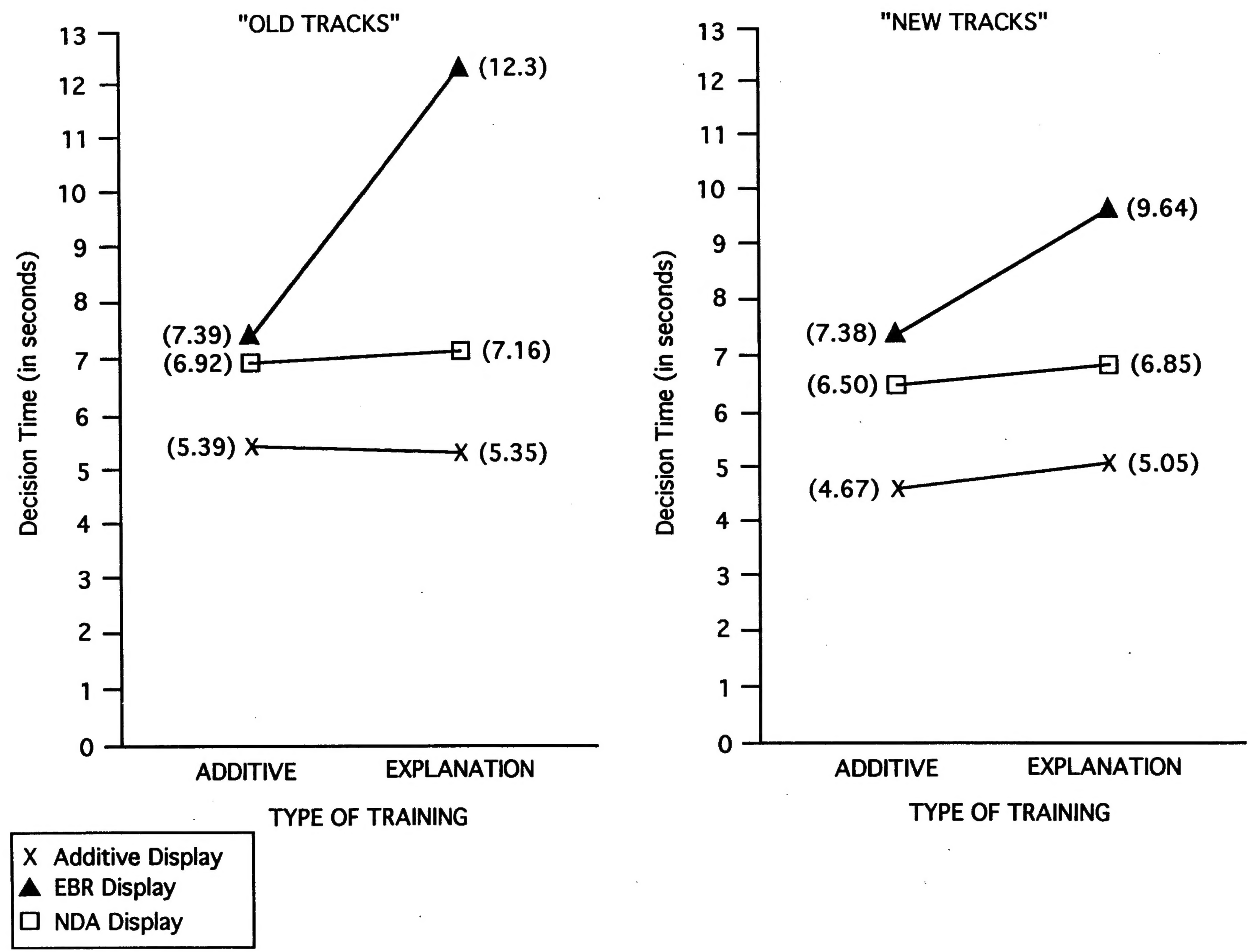


Figure 14. Mean decision time for each of the six conditions, for the two different types.